Incorporating SULF1 polymorphisms in a pretreatment CT-based radiomic model for predicting platinum resistance in ovarian cancer treatment

Xiaoping Yi [a,b,1], Yingzi Liu [c,d,e,f,1], Bolun Zhou [g,1], Wang Xiang [h], Aojian Deng [h], Yan Fu [h], Yuanzhe Zhao [h], Qianying Ouyang [c,d,e,f], Yujie Liu [c,d,e,f], Zeen Sun [c,d,e,f], Keqiang Zhang [i], Xi Li [c,d,e,f], Feiyue Zeng [a,b,*], Honghao Zhou [c,d,e,f], Bihong T. Chen [j]

[a] *Department of Radiology, Xiangya Hospital, Central South University, Changsha 410008, PR China*
[b] *National Clinical Research Center for Geriatric Disorders (Xiangya Hospital), Central South University, Changsha 410008, PR China* [c] *Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha 410008, PR China*
[d] *Institute of Clinical Pharmacology, Central South University, Hunan Key Laboratory of Pharmacogenetics, Changsha 410008, PR China* [e] *Engineering Research Center of Applied Technology of Pharmacogenomics, Ministry of Education, Changsha 410008, PR China* [f] *National Clinical Research Center for Geriatric Disorders, Changsha 410008, PR China* [g] *Xiangya School of Medicine, Central South University, Changsha 410013, PR China*
[h] *Department of Radiology, Hunan Provincial Tumor Hospital, The Affiliated Tumor Hospital of Xiangya Medical School of Central South University, Changsha 410008, PR China*
[i] *Hunan Provincial Tumor Hospital, The Affiliated Tumor Hospital of Xiangya Medical School of Central South University, Changsha 410008, PR China* [j] *Department of Diagnostic Radiology, City of Hope National Medical Center, Duarte, CA, United States*

ARTICLEINFO

Keywords:
Radiomics
Radiogenomics
Pharmacogenomics
Platinum-resistance
Ovarian cancer
Human sulfatase 1 (SULF1)

ABSTRACT

Objective: Early detection of platinum resistance for ovarian cancer treatment remains challenging. This study aims to develop a machine learning model incorporating genomic data such as Single-Nucleotide Polymorphisms (SNPs) of Human Sulfatase 1 (*SULF1*) with a CT radiomic model based on pre-treatment CT images, to predict platinum resistance for ovarian cancer (OC) treatment.

Methods: A cohort of 102 patients with pathologically confirmed OC was retrospectively enrolled into this study from January 2006 to February 2018. All patients had platinum-based chemotherapy after maximal cyto- reductive surgery. This cohort was separated into two groups according to treatment response, i.e., the group with platinum-resistant disease (PR group) and the group with platinum-sensitive disease (PS group). We genotyped 12 SNPs of *SULF1* for all OC patients using Mass Array Method. Radiomic features, SNP data and clinicopathological data of the 102 patients were used to build the differentiation models. The study participants were divided into two cohorts: the training cohort (n = 71) and the validation cohort (n = 31). Feature selection and predictive modeling were performed using least absolute shrinkage and selection operator (LASSO), Random Forest Classifier and Support Vector Machine methods. Model performance for predicting platinum resistance was assessed with respect to its calibration, discrimination, and clinical application.

Results: For prediction of platinum resistance, the approach combining the radiomics, clinicopathological data and SNP data demonstrated higher classification efficiency, with an AUC value of 0.993 (95 % CI: 0.83 to 0.98) in the training cohort and 0.967 (95 % CI: 0.83 to 0.98) in validation cohort, than the performance with only the SNPs of SULF1 model (AUC: training, 0.843 [95 %CI: 0.738-0.948]; validation, 0.815 [0.601-1.000]), or with only the radiomic model (AUC: training, 0.874 [95 %CI: 0.789-0.960]; validation, 0.832 [95 %CI: 0.687-0.976]). This integrated approach also showed good calibration and favorable clinical utility.

Conclusions: A predictive model combining pretreatment CT radiomics with genomic data such as SNPs of *SULF1* could potentially help to predict platinum resistance in ovarian cancer treatment.

* Corresponding author at: Department of Radiology, Xiangya Hospital, Central South University, Changsha 410008, PR China. *E-mail address:* zengfeiyue@csu.edu.cn (F. Zeng).
[1] These authors contribute equally.

https://doi.org/10.1016/j.biopha.2020.111013
Received 21 August 2020; Received in revised form 3 November 2020; Accepted 11 November 2020
Available online 20 November 2020
0753-3322/© 2020 The Authors. Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Ovarian cancer (OC) is the leading cause of death from gynecologic cancers in the world. Ninety percent of OC are epithelial in origin, with serous carcinoma being the most common subtype [1]. The standard treatment for OC is primary debulking surgery followed by platinum-based chemotherapy. Majority of patients respond well to initial platinum-based chemotherapy. However, an approximately 15 % of patients will develop primary resistance to platinum-based chemotherapy, and an additional 30 % will recur within 6 months after treatment has completed [2].

Pre-treatment identification of those who will not benefit from platinum-based chemotherapy (non-responder) can help guide clinical decision making. First, a priori knowledge of responsiveness to chemotherapy with radiographic evidence of a non-responder may favor an optimal individualized intervention as opposed to traditional chemotherapy strategy. Secondly, the knowledge of chemotherapy responsiveness may help to reduce the potential risk of disease progression from ineffective treatment. Lastly, identifying the risk of chemo- resistance could assist clinicians in selecting the few who may need additional time-consuming and costly testing (i.e. gene sequencing) to formulate the most effective treatment strategy utilizing the new targeted therapies.

It has been challenging to develop non-invasive methods for predicting the response to chemotherapy in OC. One strategy is to use pharmacogenomics, which is a study on

how genetic variation influences personal responses to therapy [3]. With the development of DNA sequencing, especially genome-wide association studies (GWAS) and Next Generation Sequencing (NGS) techniques, pharmacogenomics has made progress in personalized cancer treatment utilizing genetic information in recent years [4]. Many single-nucleotide polymorphism (SNP) biomarkers have been identified to predict susceptibility to drug therapy. However, clinical application of pharmacogenomics in cancer treatment has been limited.

Human sulfatase 1 (SULF1) is a heparin-degrading endosulfatase, which catalyzes the 6-O desulfation of heparan sulfate proteoglycan (HSPG). Sulfated HSPGs on the cell surface could serve as coreceptors for many growth factors, cytokines and chemokines to regulate their downstream signaling [5]. SULF1 modulates critical cellular signaling pathways through inhibiting the function of HSPGs by desulfation [6,7]. Several studies reported that SULF1 expression was associated with platinum resistance [8], and SULF1 rs2623047 GG/GA genotypes were associated with longer progression-free survival [9]. Our own previous study also found that loss of SULF1 was involved in resistance to cisplatin using the genome-scale CRISPR-Cas9 knockout (GeCKO) library [10]. We therefore speculate that SNPs of *SULF1* may be potentially useful for assessing the response to platinum-based treatment for ovarian cancer.

Radiomics is the process of extracting quantitative image features from standard-of-care medical imaging via a high-throughput analysis [11]. The core of radiomics is to extract and select valuable high-dimensional imaging features, and to combine various meaningful risk factors to build a classification model for assisting clinical decision-making such as diagnosis, prediction of treatment response, and prognosis [12]. Radiomics has been studied on a variety of cancers [13], including ovarian cancer [14], non-small cell lung cancer [15], colorectal cancer [16,17], rectal cancer [18], bladder cancer [19], and breast cancer [20]. Several recent studies have shown that radiomics can effectively predict the prognosis [21], the risk of disease progression [22] and postoperative recurrence in patients with OC [23]. However, it is not clear whether the radiomics-based strategy can effectively predict platinum resistance for patients with OC, and whether incorporation of pharmacogenomics such as SNP could further improve the classification performance of the radiomic model.

For this study, we hypothesized that radiomic features from pretreatment CT and SNPs of *SULF1* were associated with platinum resistance for OC treatment, and a machine learning radiomic model incorporating both the SNP data and pre-treatment CT texture radiomic features could more effectively predict treatment response than models built with either the SNP data or the pre-treatment CT radiomic data alone. The purpose of this study was to investigate the performance of this integrated approach for predicting platinum resistance for OC treatment. **2. Materials and methods**

2.1. Patients

The study population and data collection were described previously [24]. Briefly, a total of 102 patients with OC from Hunan Cancer Hospital were included in this retrospective study from January 2006 to February 2018 (Table 1). The study protocol was approved by the Committee for Medical Ethics, Institute of Clinical Pharmacology, Central South University (registration no. CTXY-140002-10), and was registered in the Chinese Clinical Trial Registry (ChiCTR-TNC-15007604). Informed consent was waived due to the retrospective nature of this study.

All patients had pathologically confirmed serous or mucinous epithelial OC with stage of II, III, or IV, according to the International Federation of Gynecology and Obstetrics (FIGO) criteria. All patients received standard (at least 3 cycles of) platinum-based chemotherapy within 7 days after maximal cyto-reductive surgery. Patients lacking radiological data were excluded and those with other cancers or underlying disease that may affect the treatment were also excluded.

Primary recurrence was used to determine platinum resistance and sensitivity. Platinum resistance was defined if the patient had persistent or progressive disease during treatment or recurrence within 6 months after the last cycle of chemotherapy. Platinum sensitivity was defined if the progression or recurrence occurred at six months or later after completing a platinum-based regimen. Progression free survival (PFS) was calculated as the duration, in months, from the date of initial surgical resection to the date of progression, or the date of recurrence detected by physical examination, CA125 or imaging studies. Overall survival was defined as the duration from diagnosis to death from any cause.

2.2. DNA extraction

A total of 5 mL venous blood from each patient was collected in EDTA (ethylenediaminetetraacetic acid)-containing tubes. Genomic DNA was isolated using a DNAeasy Blood & Tissue Kit (QIAGEN, Germantown, MD, USA) according to standard protocols.

2.3. SNP selection and genotyping

SNPs were identified and selected by Haploview 4.2 (Broad Institute, Cambridge, MA, USA) using pair-wise tagging with default settings (pairwise r2 threshold ≥ 0.8). The following criteria were used to select relevant SNPs: 1) selecting the SNPs with minor allele frequency (MAF) ≥ 5% for the Chinese population at the Southern region according to the 1000 Genomes database; 2) selecting the SNPs located in the promoter region, in the 5′- untranslated region (5′-UTR), coding regions, or in the 3′-UTRs; 3) selecting the SNPs that were reported to be clinically relevant according to the previous literature. If the SNPs were in high linkage disequilibrium (LD) (r 2 ≥ 0.8), only one SNP was genotyped. As a result, 12 SNPs were selected for genotyping and analysis (Table 2). Sequenom Mass Array system was used for genotyping.

2.4. CT imaging

All patients underwent a routine medically-indicated CT scan on a 16-MDCT (Brilliance 16, Philipps) or a 64-MDCT (SOMATOM

Table 1
Demographic, clinicopathological and laboratory findings of 102 patients with EOC.

Characteristic	Total (n = 102)	PR group (n = 26)	PS group (n = 76)	*P* Value	Training Cohort (n = 71)	Validation Cohort (n = 31)	*P* Value
Demographics and clinical characteristics							
Age (y)#	52.2 ± 8.4	55.1 ± 7.2	51.2 ± 8.6	0.352	52.7 ± 8.5	51.0 ± 7.8	0.042
Female	155.0 (140.0− 166.0)	155.0 (141.0− 166.0)	155.0 (140.0− 164.0)	0.582	155.0(140.0− 166.0)	155.0(142.0− 161.0)	0.729
Height	55.7 ± 7.4	55.2 ± 6.2	56.0 ± 7.8	0.662	55.5 ± 7.8	56.2 ± 6.3	0.691
Weight	36.5(35.2− 37.0)	36.5(35.2− 36.9)	36.6(35.5− 37.0)	0.100	36.5(35.2− 37.0)	36.5(35.5− 36.9)	0.962
Temp	122.2 ± 13.3	120.3 ± 14.5	123.0 ± 13.0	0.684	122.6 ± 14.3	121.4 ± 10.9	0.385
SP	77.0 ± 9.6	76.5 ± 10.4	77.1 ± 9.3	0.739	76.8 ± 10.0	77.5 ± 8.5	0.755
DP	43.0(25.0− 83.0)	40.5(25.0− 65.0)	44.0(28.0− 83.0)	0.396	45.0(25.0− 83.0)	42.0(29.0− 69.0)	0.400
SBP-DBP	23.2 ± 2.7	23.1 ± 2.0	23.2 ± 2.9	0.635	22.1 ± 2.6	22.9 ± 2.7	0.858
BMI							
Laboratory findings	7.0 ± 2.2	7.2 ± 2.2	6.9 ± 2.2	0.312	7.1 ± 2.3	6.6 ± 1.9	0.475
White blood cell count, × 10^9 per L	4.7 ± 2.1	4.9 ± 2.1	4.7 ± 2.0	0.437	4.8 ± 2.2	4.5 ± 1.8	0.628
NEU, × 10^9 per L	66.3(13.2− 92.4)	66.7(13.2− 83.1)	66.1(37.6− 92.4)	0.884	67.5(14.6− 84.31)	64.4(13.2− 92.4)	0.430
NEU%	1.6(0.2− 5.5)	1.6(0.8− 2.5)	1.6(0.2− 5.5)	0.869	1.5(0.2− 5.5)	1.7(0.8− 2.6)	0.363
Lymphocyte count,× 10^9 per L	23.5(7.5− 71.8)	19.8(7.5− 71.8)	24.6(7.5− 56.3)	0.249	22.5(7.5− 52.3)	26.3(7.5− 71.8)	0.298
LYM%	0.5(0− 1.0)	0.6(0.3− 1.0)	0.4(0− 1.0)	0.016	0.5(0.2− 1.0)	0.4(0− 0.8)	0.440
MONO, × 10^9 per L	7.4 ± 2.5	8.3 ± 1.8	0.5 ± 0.2	0.644	7.5 ± 1.9	7.2 ± 3.4	0.014
MONO%	0.1(0− 6.1)	0.1(0− 1.1)	0.1(0− 6.1)	0.860	0.1(0− 0.7)	0.1(0− 6.1)	0.785
EOS, × 10^9 per L	1.1(0.1− 14.1)	1.1(0.1− 14.1)	1.2(0.1− 8.5)	0.985	1.0(0.1− 8.5)	1.2(0.1− 14.1)	0.460
EOS%	58.7 ± 18.5	63.2 ± 14.4	57.2 ± 20.0	0.366	59.8 ± 17.5	56.2 ± 20.9	0.156
TSGF(U/ML)	9.2(1.4− 756.0)	9.8(2.4− 201.0)	9.0(1.4− 756.0)	0.972	8.6(1.4− 309.1)	9.7(2.3− 756.0)	0.833
CA199(U/ML)CA	6.9(1.2− 24.4)	9.1(1.2− 24.4)	6.6(1.5− 24.4)	0.090	7.4(1.2− 24.4)	6.2(1.9− 21.8)	0.136
NSE(ng/mL)	0.9(0.2− 34.2)	1.0(0.4− 9.1)	0.9(0.2− 34.2)	0.167	1.0(0.3− 34.2)	0.8(0.2− 7.4)	0.250
CEA							
CEA (+, >5 ng/mL)	5.0(0.7− 222.4)	4.5(1.2− 50.0)	5.0(0.7− 222.4)	0.721	4.0(1.2− 222.4)	5.6(0.7− 125.4)	0.413
CA242	1.4(0.3− 10.0)	1.4(0.6− 5.5)	1.4(0.3− 10.0)	0.202	1.4(0.3− 6.7)	1.3(0.3− 10.0)	0.590
AFP	410.9(5.0− 4524.0)	421.8(6.6− 3818.0)	410.9(5.0− 4524.0)	0.902	399.0(6.6− 4524.0)	533.1(5.0− 1301.1)	0.356
CA125(U/ML)	13.6(0.2− 244.7)	16.1(2.8− 121.9)	12.8(0.2− 244.7)	0.253	14.4(0.2− 244.7)	11.4(2.6− 225.5)	0.631
CA153	5.7(3.3− 14.7)	5.6(3.3− 6.3)	5.7(4.3− 14.7)	0.338	5.6(3.3− 14.7)	5.7(4.8− 6.3)	0.418
Glycosylated hemoglobin							
Pathological findings							
Stage FIGO (n, %)				0.028*			0.794
II	8 (7.8)	8 (10.5)	0 (0)		6 (8.5)	2 (6.5)	
III	80 (78.5)	61 (80.3)	19 (73.1)		54 (76.1)	26 (83.9)	
IV	14 (13.7)	7 (9.2)	7 (26.9)		11 (15.5)	3 (9.7)	
Grade (differentiation)				0.074			0.263
Low	74 (72.5 %)	26(100.0 %)	65(85.5 %)		61(85.9 %)	30(96.8 %)	
Moderate	8 (7.8 %)	0	10(13.2 %)		9(12.7 %)	1(3.2 %)	
High	1 (1.0 %)	0	1(1.3 %)		1(1.4 %)	0	

Note.—Unless otherwise indicated, data are numbers of patients, and data in parentheses are percentages. # age was presented as media (minimum~ maximum). * P < .05 suggests a significant difference between the age in the two cohorts. COPD, chronic obstructive pulmonary disease.

Table 2 twelve selected SNPs in SULF1.

SNP (rs number)	Allele	Region	Call rate	MAF	HWE
rs2623047	A > G	promoter	100	0.423	0.132
rs59465016	G > C	promoter	100	0.474	0.629
rs10957496	A > C	intron	100	0.451	0.252
rs1372150	G > A	intron	97.9	0.289	0.675
rs1441199	C > T	intron	100	0.097	0.086
rs16936018	C > T	intron	100	0.441	2.084
rs16936195	G > A	intron	100	0.187	0.764
rs1899274	T > A	intron	100	0.228	1.296

rs2028442	A > G	intron	100	0.210	0.540
rs2583092	G > A	intron	100	0.100	0.143
rs4737999	A > G	intron	100	0.285	1.057
rs3802278	G > A	3′UTR	100	0.362	1.603

MAF: Minor allele frequency; HWE: Hardy-Weinberg equilibrium.

Definition, Siemens) scanner. After a non-contrast scan, a contrast- enhanced scan was performed using a power injector (Ulrich CT plus 150, Ulrich Medical, Ulm, Germany). Briefly, 90− 100 ml of iodinated contrast material (Ultravist 370, Bayer Schering Pharma, Berlin, Germany) was injected intravenously at a rate of 3.0–3.5 ml/s. Enhanced images at portal-venous phase (scan with fixed delay time of 60− 70 s) were available for all patients. all patients had both the unenhanced and the portal-venous phase images and CT images from these two phases were analyzed. All CT images were retrieved from the Picture Archiving and Communication Systems (PACS, Carestream, Canada), and were transferred to an external workstation (Leonardo; Siemens Medical Solutions, Forchheim, Germany). All CT images were subsequently reconstructed with a thickness of 5 mm.

2.5. Imaging analysis

CT images for each patient were reviewed independently by two abdominal radiologists (reader 1. WX with >8 years of experience; reader 2, XY with 10 years of experience) respectively. Radiologists were blinded to the patients' information such as the radiological and clinicopathological data. All CT imaging findings of the ovarian tumors were recorded including: longitudinal diameter (mm), antero-posterior diameter (mm), axial diameter (mm), pre-enhanced CT value (CTpre) (Hu), enhanced CT value (CTpost) (Hu), enhancement range (Hu). Any discrepancy between the two readers was reached by consensus through discussion. Image segmentation and textural features extraction were performed using MaZda software (Version 4.6, Institute of Electronics, Technical University of Lodz, Poland) for three-dimensional manual segmentation of the region of interest (ROI) according to the manual operation instructions. Briefly, based on enhanced CT images of portal- venous phase, the entire tumor was manually segmented by the study radiologists (WX and XY). Then, 340 radiomic features based on the CT images were automatically extracted.

2.6. Statistical analysis, feature selection, signature generation, and prediction model building

The SPSS 20.0 (SPSS Inc, Chicago, IL) and Plink 1.07 were used for general statistics analyses. The Hardy-Weinberg equilibrium (HWE) of SNPs among all subjects was examined via a goodness-of-fit χ2-test to compare the expected genotype frequencies with those that were observed. To test the differences between groups, the fisher's exact test or the chi-square test was performed for the qualitative imaging features and the Wilcoxon rank-sum test was performed for the quantitative imaging features.

For radiomic data, we used MATLAB 2017a (The Mathworks, Inc., Natick, MA, USA) to perform all data processing, data reduction and feature selection, and model building. The most useful sets of predictive features were selected from the primary data via the least absolute shrinkage and selection operator (LASSO) method, and a feature signature was calculated for each patient at the same time as a linear combination of selected features weighted by their respective coefficients. Three types of signatures, named Rad- or clinicopathology- or SNP-score were generated based on radiomic features, clinicopathological features or SNP data, respectively.

An additional classification model was constructed by Random Forest (RF) based on corresponding selected features, and a RF-score was generated. Subsequently, the support vector machine (SVM) method (SVM-score) was used to build a combined classification model, based on Rad-score and RF-score in the previous steps. A final classification model was generated based on the three SVM-scores that were calculated from radiomic texture features, clinicopathological features and SNP features respectively and was obtained using the SVM method again (SVM-score-final). The three types of data (radiomic texture features, clinicopathological data and SNP data) were used to build the final set of models. A total of ten models were generated representing each classification task through the above steps. These ten models consisted of three models generated from the radiomic texture features, three models generated from clinicopathological features, three models generated from SNP features, and one combined SVM model.

The basis of this algorithm was to regard LASSO and RF as weak regressors and to combine them via SVM. For each type of data, we first used LASSO to obtain the Rad-score, and then used its side product as the feature set to obtain the RF-score. After Rad-score and RF-score were respectively acquired by two different weak regressors, we then used SVM to regress them in a two-dimensional plane to achieve a better result than by itself without SVM. Furthermore, data sets from the traditional radiological-clinicopathological data and radiomic texture features were independent to each other. Therefore, for the same reason, we used SVM to regress the scores from the traditional radiological- clinicopathological data and from the radiomic texture features to get the final regression score. Finally, the regressed scores could be binarized for further prediction.

The models were developed in the training cohort and tested in the validation cohort. The classification efficiencies such as the receiver operating characteristic (ROC) curves for each model mentioned above were calculated for both the training cohort and validation cohort. Details of the work flow for building the classification model were shown in Fig. 1. A P value < 0.05 was considered statistically significant.

2.7. Performance, validation and clinical application of the nomogram

Calibration curves (Hosmer-Lemeshow H test) were used to evaluate the calibration of the models, and ROC curves were performed to access the differentiation efficiency. The clinical usefulness of the models was evaluated by using decision curve analysis in the validation cohort. *Additionally, a correlation matrix analysis was performed to evaluate the correlations among all selected features that were integrated into the final predictive model.*

3. Results

A total of 102 patients with OC were retrospectively enrolled into this study. Details of the exclusion criteria and the patient recruiting process are shown in Fig. 2. For the radiomic analysis, all included cases were randomly separated into a training cohort and a validation cohort at a ratio of 7:3.

The demographic characteristics of the patients with OC were summarized in Table 1. Among these patients, 76 (74.5 %) were in the PS group and 26 (25.5 %) were in the PR group. No difference was found in age, histologic subtype and tumor differentiation type between the PS group and the PR group, except for FIGO stage (P = 0.028). Information of all the tested SNPs was presented in Table 1.

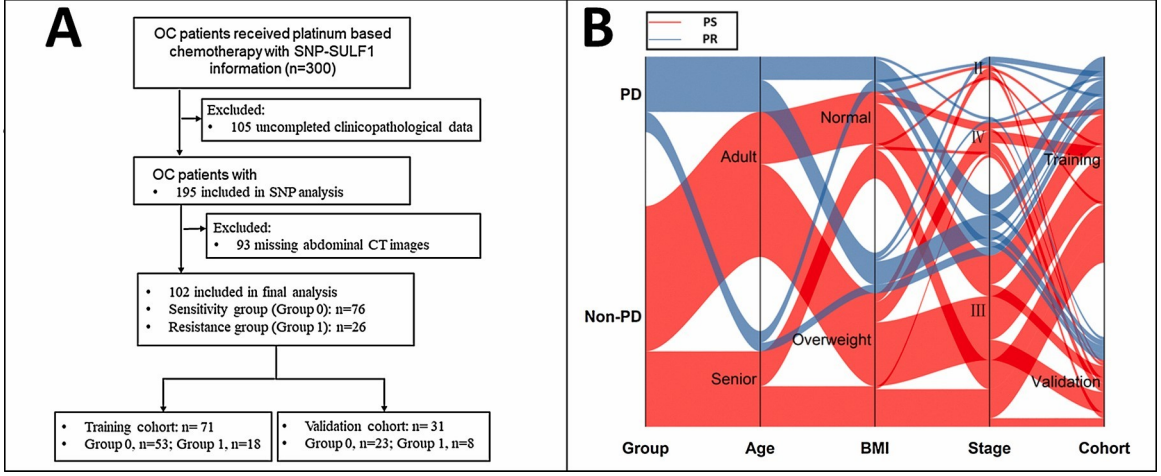
3.1. Inter-observer and intra-observer reproducibility of radiomic features

Our results demonstrated satisfactory inter- and intra-observer reproducibility of the radiomic features extraction. The inter-observer intraclass correlation coefficient (ICCs) of radiomic features between reader 1 and reader 2 ranged from 0.761 to 0.893. The intra-observer ICC of reader 1's two extraction performances ranged from 0.781 to 0.909. As a result, the radiomics features extracted by reader 1 were used in all subsequent calculations.

3.2. Classification model building and model performance

We have extracted a total of 340 radiomic features from enhanced abdominal CT images for each patient. In addition, we obtained 34 clinicopathological features and 28 pieces of SNP data. Utilizing a combination of the radiomic, clinicopathological and SNP data, a set of features were selected by LASSO. Based on the scores (weak regressors) fitted by the selected data for the radiomics, clinicopathological and SNP features respectively, a predictive model was finally constructed with the SVM method for non-responder prediction by using modeling strategy described in our previous study [28].

Fig. 3 (A-D) demonstrated the performance data of the



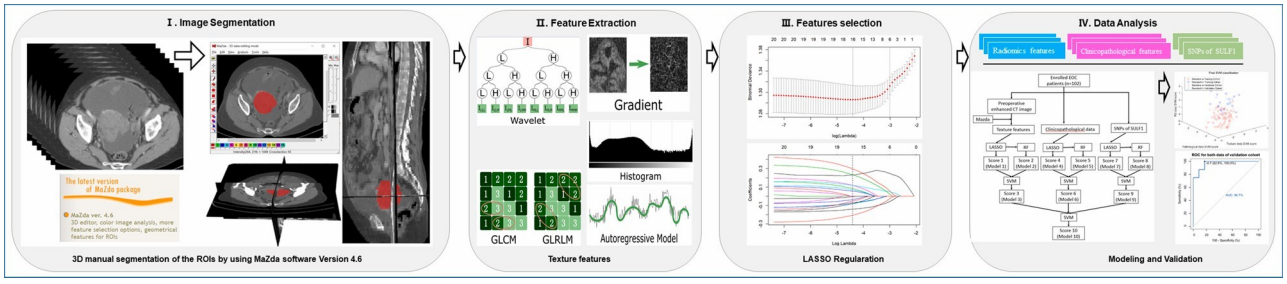


Fig. 1. Workflow of extracting radiomic data from CT images, radiomic feature reduction and modeling. (I)Representative tumor segmentation on CT images. (II) Radiomic feature extraction. (III) Radiomic feature selection. (IV) Predictive modeling with machine learning approach.
Fig. 2. (A) Flow-chart presenting the recruiting process and the exclusion criteria. (B) Sankey diagram showing the scmap cluster projection for treatment response, gender and age composition, and pathological stage.
Note: Senior: >60 years old, Stage: Pathological stage, OC: ovarian cancer.

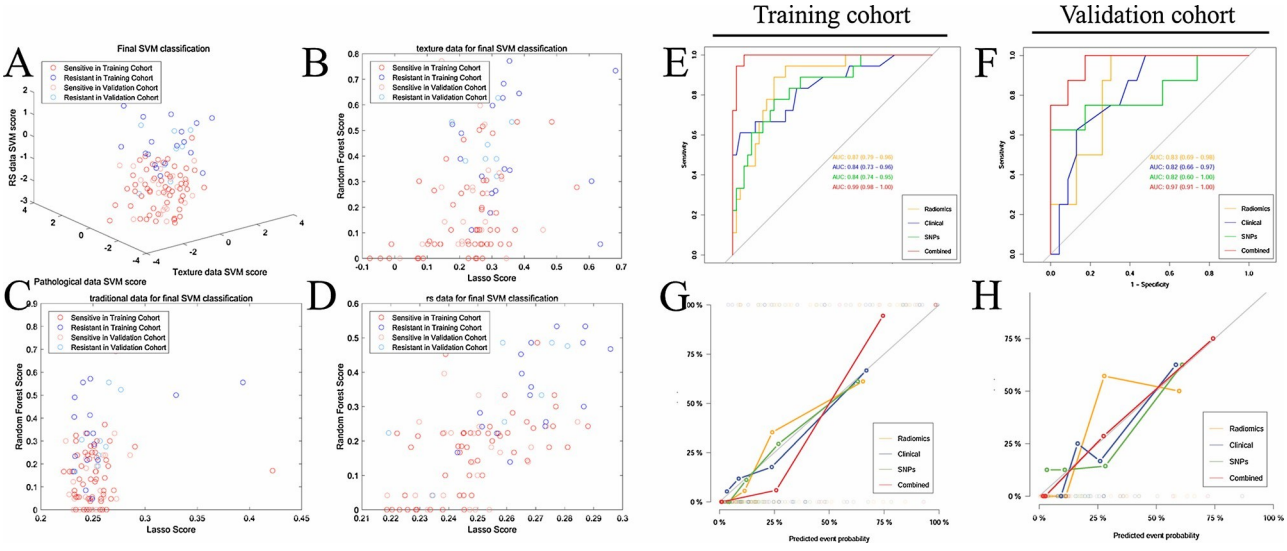


Fig. 3. Classification results presented by case distribution map. A. Combined model, B. Radiomic model, C. Clinical model and D. SNP model. Receiver operating characteristic (ROC) curves of these four models in the training cohort (E) and the validation cohort (F) were presented. The calibration curves calculated in training
cohort (G) and validation cohort (H) were presented.

differentiation models. Based on the selected radiomic features, clinicopathological data and SNP features, a predictive model was finally constructed with the SVM method for non-responder prediction. The final SVM model yielded an AUC of 99.3 % (95 %CI: 98.0–100.0 %) in the training cohort (with a sensitivity of 100.0 % [95 %CI: 94.4–100.0 %], and a specificity of 96.2 % [95 %CI: 90.6–100.0 %]), and an AUC of 96.7 % (95 %CI: 91.3–100.0 %) (with a sensitivity of 100.0 % [95 %CI: 75.0–100.0 %] and a specificity of 91.3 % [95 %CI: 73.9–100.0 %]) in the validation cohort (Fig. 3, E, F). The calibration curve for the probability of resistance in the training and validation cohorts showed good agreement between prediction and observation (Fig. 3 G, H).

Fig. 4 showed the correlation matrix maps for the three types of features included in the prediction models, i.e., radiomic, clinicopathological and SNP features. Considering that these features were independently selected after performing feature selection with the LASSO method, the correlation values in Fig. 5B could be used to assess whether

these features contributed to the classification performance in a similar way showing strong correlations or from a different perspective showing weak correlations. Our results showed weak correlations between SNP and other two types of features (radiomic and clinicopathological features) as indicated by the low correlation scores (all <0.400). These findings suggested that these three types of features may represent or reflect tumor characteristics from a different perspective and therefore integration of these features into the final prediction models may enhance the classification performance.

3.3. Clinical application

The decision curve analysis for the differentiation model was presented in Fig. 5. The decision curve showed that if the threshold probability of a patient or doctor was greater than 5%, using the combined model to predict resistance in OC patients would add more benefit than either the diagnose-all-patients scheme or the diagnose-none scheme. Moreover, use of combined model demonstrated significant clinical benefits than any model built with a single data source such as radiomics, clinicopathological or SNPs data) (Fig. 5).

4. Discussion

In this study, we found that incorporating SNPs of SULF1 into a pretreatment CT- based radiomic model greatly improved the performance of the model for predicting platinum resistance in patients with OC. Through a radiomic approach with machine learning method, the final model combining the genomic SNP data with the radiomic data

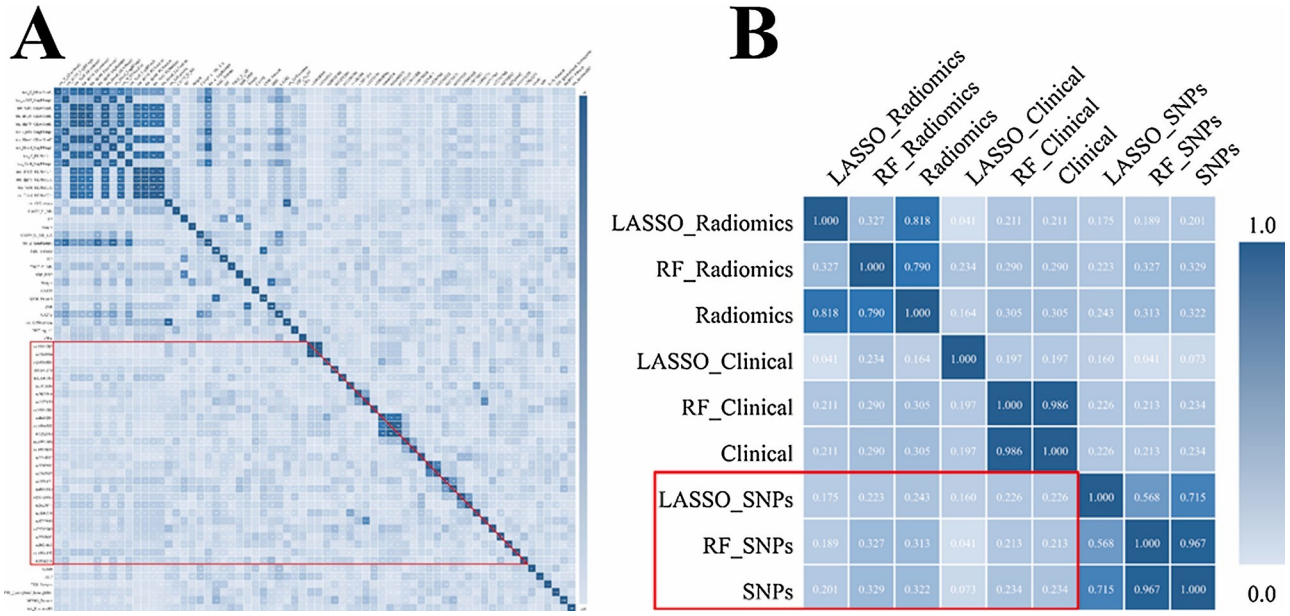


Fig. 4. Correlation matrix map. (A) Correlation matrix map for features included in the final prediction models. The area marked by a red frame refers to the correlation coefficient diagram between Single-Nucleotide Polymorphism (SNP) features and the other two types of features (radiomic features and clinicopathological features). (B) Correlation matrix map and correlation scores generated during modeling with different types of features (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

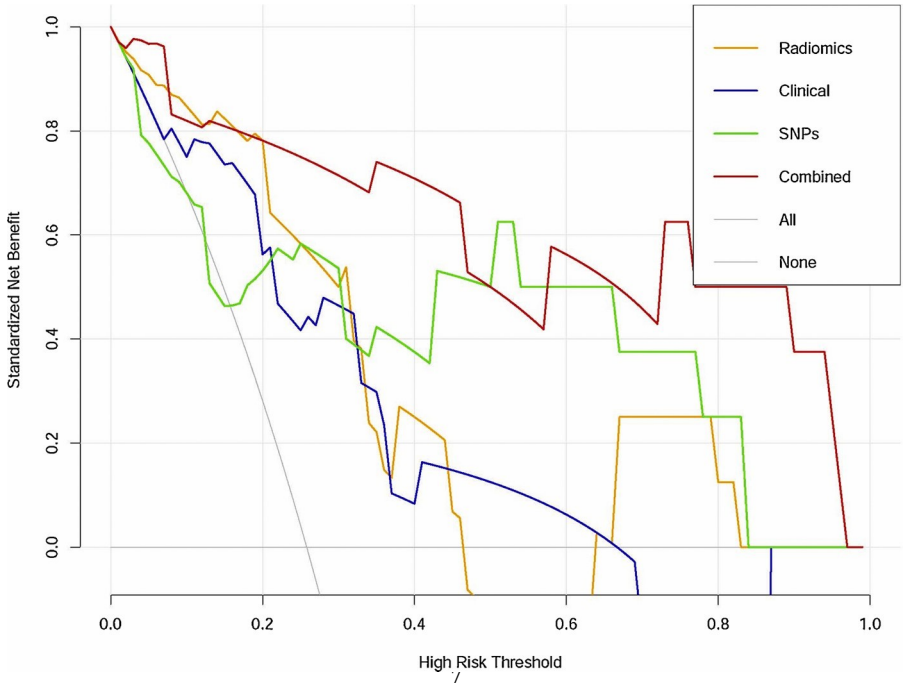


Fig. 5. Decision curve analysis for the models built with SNPs, radiomics, clinicopathological and a combination of models respectively. The y-axis measures the net benefit. Each of the different colour line represented these four types of models, respectively. The horizontal grey line represented the assumption that no patients had risk of resistant disease and the oblique grey line represented all patients would develop chemotherapy resistance.

achieved a better discrimination performance than using the SNP data or the CT radiomic data alone.

Radiomics has been shown as a high-performance method for accurately predicting treatment response through assessment of tumor heterogeneity [25]. Tumor heterogeneity has been observed at the histological and genetic levels, and has been reported to be associated with adverse clinical outcomes [26,27]. Image-based methods are more feasible and are readily available to comprehensively evaluate tumor heterogeneity with the capacity to assess the entire tumor. It is not surprising that radiomics could be used to predict therapeutic response as shown in our previous study [28]. Prior report has shown that radiomic methods could be used to evaluate outcomes of patients with OC in a non-invasive and effective approach [21,29]. In the present study, we found that pretreatment CT radiomics could indeed be potentially useful for identifying patients with chemotherapy resistance prior to surgery and our predictive models showed a satisfying performance.

We understand that treatment response might be determined by various factors. SNP genotyping may only reflect part of the reason for platinum resistance in OC treatment, and a single positive SNP as a biomarker alone may not be sufficient to predict treatment response. Moreover, different studies have identified inconsistent SNP biomarkers [4]. Therefore, developing a prediction model for treatment response through the combined use of SNP and other potential biomarkers is needed. In this study, we used SNP information as an independent data source and we combined it with both the clinicopathological data and imaging features to build a predictive model, which should be more effective in predicting the risk of platinum resistance in ovarian cancer.

Our predictive model incorporating the SNP data into CT radiomic model achieved a higher efficiency for predicting platinum resistance than using either the SNP data or the CT radiomics alone. *Our results showed that the novel integrated factor (SNPs of SULF1) contributed greatly to the differentiation model, indicating the potential value of SULF1 for assessing platinum resistance for OC treatment.*

In terms of mechanism of action of SULF1, it has been shown that SULF1 modulates critical cellular signaling pathways through inhibition of HSPG function by desulfation [6,7]. Dysregulation of SULF1 may result in altered pathways and cause changes in the tumor cell microenvironment which is critical for tumor sensitivity to platinum-based chemotherapy [8]. Therefore, our results were not unexpected in that incorporation of SULF1 polymorphisms into a radiomic model helped to predict platinum resistance in ovarian cancer treatment. Our data also suggested that SNPs of SULF1 may reflect different inherent characteristics of OC for platinum resistance than what the radiomic features may imply. Furthermore, it was not unreasonable to speculate that modification of tumor microenvironment may potentially lead to macroscopic changes in tumor heterogeneity, which could be detected on imaging as radiomic features. Our data and its implications supported the approach of incorporating SNPs of SULF1 into the radiomic model to enhance the performance of our prediction models.

Our data showed that the clinicopathological characteristics were essential for the classification efficiency of our predictive model. For example, body height has been a key factor for the current dose calculation of chemotherapy drugs. A recent study has showed that the poor prognosis of malignancies is significantly associated with height of patients [30]. Abnormal body mass index (BMI) in cancer patients at diagnosis has been associated with lower survival rates. More interestingly, among patients with Ewing sarcoma who underwent induction chemotherapy, higher BMI was significantly associated with worse histologic response rate and worse prognosis [31]. Survival was influenced by baseline BMI status in patients without metastatic disease with a locally advanced esophageal cancer treated with definitive chemoradiotherapy [32]. In addition, Laboratory values such as absolute counts of lymphocytes and platelets, platelet/lymphocyte ratio (PLR), platelet/neutrophil ratio, platelet/monocyte ratio (all $p < 0.001$), and neutrophil/lymphocyte ratio (NLR) ($P = 0.013$) have also shown statistically significant differences among three different response subgroups (responsive, stable, and non-responsive) in patients with OC receiving dose-dense platinum based chemotherapy [33].

Additionally, the baseline expression of tumor markers or their trend of changes after treatment, such as CA12-5 and CA24-2, are important indicators for evaluating the chemotherapy response and prognosis in patients with OC [34]. Previous study has shown that a CA-125 decrease following each cycle indicated a response in patients with recurrent OC undergoing chemotherapy [35]. Pathological grade is also closely related to chemotherapy response and prognosis in patients with OC [36]. Similarly, the intrinsic characteristics of tumors reflected by these clinicopathological markers, radiomic features or SNPs of SULF1 markers may be very different, according to the correlation matrix analysis. Therefore, addition of these markers could improve the performance of the models for predicting treatment response as shown in our current study.

Our study approach had some similarities but also some advantages over the published reports using pre-treatment imaging data and machine learning to evaluate treatment response in cancer research [28, 37–39]. For example, previous studies tend to select only a small number of features to achieve the desired goal and to avoid overfitting their prediction models. This approach may have inadvertently removed many potentially useful features resulting in decreased accuracy of the predictive model. In our study, we avoided this issue of losing useful features by integrating valuable features into a single factor, and then entering this newly created factor as a single feature (weak classifier) into a classic machine learning method such as SVM for modeling. The valuable features were selected by using reasonable feature dimensionality reduction and screening methods with LASSO and RF, as performed in our own previously published report [28]. It should be noted that the usual drawback of machine learning, that is, the non-interpretability of the differentiation model, also existed in this research method [40].

Our study had several limitations. First, selection bias may exist given the retrospective nature of this study. Second, the sample size of our study was modest, which may potentially affect the stability and reliability of our model. Third, this study was limited as both the training cohort and the validation cohort were from the same institution. External validation with an independent cohort from other institutions will be necessary to validate our model performance. Future prospective multicenter study should be conducted.

5. Conclusion

In summary, our study showed that incorporating SNPs of SULF1 into a pretreatment CT radiomic model could achieve a promising performance for predicting platinum resistance for OC treatment. Future prospective multicenter studies with a

larger sample size will be needed to validate our study result and to optimize the prediction models for clinical practice.

Declaration of Competing

Interest None.

Acknowledgements

This work was supported by National Natural Science Foundation of China (81974512, 81402968), and Hunan Natural Science Foundation (2018JJ2641, 2020JJ4894). We would like to thank Dr. Guangwu Lei (Xiangya Hospital, Changsha, PR China) and Dr. Peipei Pang (GE HealthCare, Hangzhou, PR China) for their assistance in data analysis.

References

[1] L.A. Torre, B. Trabert, C.E. DeSantis, K.D. Miller, G. Samimi, C.D. Runowicz, M.
M. Gaudet, A. Jemal, Siegel RL, Ovarian cancer statistics, 2018, CA Cancer J. Clin. 68 (4) (2018) 284–296.
[2] E.L. Christie, D.D.L. Bowtell, Acquired chemotherapy resistance in ovarian cancer, Ann. Oncol. 28 (suppl_8) (2017) viii13-viii15.
[3] H.E. Wheeler, M.L. Maitland, M.E. Dolan, N.J. Cox, Ratain MJ, Cancer pharmacogenomics: strategies and challenges, Nat. Rev. Genet. 14 (1) (2013)
23–34.
[4] R.M. Weinshilboum, L. Wang, Pharmacogenomics: precision medicine and drug response, Mayo Clin. Proc. 92 (11) (2017) 1711–1722.
[5] X. He, A. Khurana, D. Roy, S. Kaufmann, V. Shridhar, Loss of HSulf-1 expression enhances tumorigenicity by inhibiting Bim expression in ovarian cancer, Int. J. Cancer 135 (8) (2014) 1783–1789.
[6] S.D. Rosen, H. Lemjabbar-Alaoui, Sulf-2: an extracellular modulator of cell signaling and a cancer target candidate, Expert Opin. Ther. Targets 14 (9) (2010) 935–949.
[7] S. Sarrazin, W.C. Lamanna, J.D. Esko, Heparan sulfate proteoglycans, Cold Spring Harb. Perspect. Biol. 3 (7) (2011).
[8] J. Staub, J. Chien, Y. Pan, X. Qian, K. Narita, G. Aletti, M. Scheerer, L.R. Roberts, J. Molina, V. Shridhar, Epigenetic silencing of HSulf-1 in ovarian cancer:
implications in chemoresistance, Oncogene 26 (34) (2007) 4969–4978. [9] C.H. Han, Y.J. Huang, K.H. Lu, Z. Liu, G.B. Mills, Q. Wei, L.E. Wang, Polymorphisms in the SULF1 gene are associated with early age of onset and survival of ovarian cancer, J. Exp. Clin. Cancer Res. 30 (1) (2011) 5.
[10] Q. Ouyang, Y. Liu, J. Tan, J. Li, D. Yang, F. Zeng, W. Huang, Y. Kong, Z. Liu, H. Zhou, et al., Loss of ZNF587B and SULF1 contributed to cisplatin resistance in ovarian cancer cell lines based on Genome-scale CRISPR/Cas9 screening, Am. J.
Cancer Res. 9 (5) (2019) 988–998.
[11] J.W. Beall, R.M. DeHart, R.M. Riggs, J. Hensley, perceived stress, stressors, and coping mechanisms among doctor of pharmacy students, Pharmacy Basel (Basel) 3
(4) (2015) 344–354.
[12] P. Lambin, R.T.H. Leijenaar, T.M. Deist, J. Peerlings, E.E.C. de Jong, J. van Timmeren, S. Sanduleanu, R. Larue, A.J.G. Even, A. Jochems, et al., Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (12) (2017) 749–762.
[13] V. Verma, C.B. Simone 2nd, S. Krishnan, S.H. Lin, J. Yang, S.M. Hahn, the rise of radiomics and implications for oncologic management, J. Natl. Cancer Inst. 109 (7) (2017).
[14] S. Nougaret, M. Tardieu, H.A. Vargas, C. Reinhold, S. Vande Perre, N. Bonanno, E. Sala, I. Thomassin-Naggara, Ovarian cancer: an update on imaging in the era of radiomics, Diagn. Interv. Imaging (2018).
[15] Y. Huang, Z. Liu, L. He, X. Chen, D. Pan, Z. Ma, C. Liang, J. Tian, C. Liang, Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung Cancer, Radiology 281 (3)
(2016) 947–957.
[16] Y.Q. Huang, C.H. Liang, L. He, J. Tian, C.H.S. Liang, X. Chen, Z.L. Ma, Z.Y. Liu, Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal Cancer, J. Clin. Oncol. 34 (18) (2016)
2157–2164.
[17] C. Liang, Y. Huang, L. He, X. Chen, Z. Ma, D. Dong, J. Tian, C. Liang, Liu Z, The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage III-IV colorectal cancer, Oncotarget 7 (21)
(2016) 31401–31412.
[18] Z. Liu, X.Y. Zhang, Y.J. Shi, L. Wang, H.T. Zhu, Z. Tang, S. Wang, X.T. Li, J. Tian, Sun YS, Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal Cancer, Clin. Cancer Res. 23 (23) (2017) 7253–7262.
[19] S. Wu, J. Zheng, Y. Li, H. Yu, S. Shi, W. Xie, H. Liu, Y. Su, J. Huang, T. Lin, A Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer, Clin. Cancer Res. 23 (22) (2017) 6904–6911.
[20] F. Valdora, N. Houssami, F. Rossi, M. Calabrese, A.S. Tagliafico, Rapid review:
radiomics and breast cancer, Breast Cancer Res. Treat. 169 (2) (2018) 217–229.
[21] H.A. Vargas, H. Veeraraghavan, M. Micco, S. Nougaret, Y. Lakhman, A.A. Meier, R. Sosa, R.A. Soslow, D.A. Levine, B. Weigelt, et al., A novel representation of inter- site tumour heterogeneity from pre-treatment computed tomography textures classifies ovarian cancers by clinical outcome, Eur. Radiol. 27 (9) (2017)
3991–4001.
[22] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, V. Buscarino, A. Colarieti, F. Tomao, G. Aletti, V. Zanagnolo, M. Del Grande, et al., Radiomics of high-grade serous ovarian cancer: association between quantitative CT features, residual tumour and disease progression within 12 months, Eur. Radiol. 28 (11) (2018) 4849–4859. [23] W. Wei, Y. Rong, Z. Liu, B. Zhou, Z. Tang, S. Wang, D. Dong, Y. Zang, Y. Guo, J. Tian, Radiomics: a novel CT-based method of predicting postoperative recurrence in ovarian cancer, Conference Proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Annual Conference 2018 (2018)
4130–4133.
[24] T. Li, J. Peng, F. Zeng, K. Zhang, J. Liu, X. Li, Q. Ouyang, G. Wang, L. Wang, Z. Liu, et al., Association between polymorphisms in CTR1, CTR2, ATP7A, and ATP7B and platinum resistance in epithelial ovarian cancer, Int. J. Clin. Pharmacol. Ther. 55
(10) (2017) 774–780.
[25] E. Sala, E. Mema, Y. Himoto, H. Veeraraghavan, J.D. Brenton, A. Snyder, B. Weigelt, H.A. Vargas, Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging, Clin. Radiol. 72 (1)
(2017) 3–10.
[26] E.C. de Bruin, N. McGranahan, R. Mitter, M. Salm, D.C. Wedge, L. Yates, M. Jamal- Hanjani, S. Shafi, N. Murugaesu, A.J. Rowan, et al., Spatial and temporal diversity in genomic instability processes defines lung cancer evolution, Science 346 (6206)
(2014) 251–256.
[27] N. Andor, T.A. Graham, M. Jansen, L.C. Xia, C.A. Aktipis, C. Petritsch, H.P. Ji, C. C. Maley, Pan-cancer analysis of the extent and consequences of intratumor heterogeneity, Nat. Med. 22 (1) (2016) 105–113.
[28] X. Yi, Q. Pei, Y. Zhang, H. Zhu, Z. Wang, C. Chen, Q. Li, X. Long, F. Tan, Z. Zhou, et al., MRI-based radiomics predicts tumor response to neoadjuvant chemoradiotherapy in locally advanced rectal Cancer, Front. Oncol. 9 (2019) 552.
[29] H. Zhang, Y. Mao, X. Chen, G. Wu, X. Liu, P. Zhang, Y. Bai, P. Lu, W. Yao, Y. Wang, et al., Magnetic resonance imaging radiomics in categorizing ovarian masses and predicting clinical outcome: a preliminary study, Eur. Radiol. 29 (7) (2019)
3358–3371.
[30] B.C. Fu, M. Song, X. Li, J. Han, H.O. Adami, E.L. Giovannucci, Mucci LA, Height as a mediator of sex differences in cancer risk, Ann. Oncol. 31 (5) (2020) 634–640.
[31] G. Goldstein, E. Shemesh, T. Frenkel, J.M. Jacobson, A. Toren, Abnormal body mass index at diagnosis in patients with Ewing sarcoma is associated with inferior tumor necrosis, Pediatr. Blood Cancer 62 (11) (2015) 1892–1896.
[32] F. Di Fiore, S. Lecleire, D. Pop, O. Rigal, H. Hamidou, B. Paillot, P. Ducrott´e, E. Lerebours, P. Michel, Baseline nutritional status is predictive of response to treatment and survival in patients treated by definitive chemoradiotherapy for a locally advanced esophageal cancer, Am. J. Gastroenterol. 102 (11) (2007)
2557–2563.
[33] C.Y. Huang, Y.C. Yang, K.L. Wang, T.C. Chen, J.R. Chen, C.S. Weng, H.J. Chien, C. L. Chang, Possible surrogate marker for an effective dose-dense chemotherapy in treating ovarian cancer, Taiwan. J. Obstet. Gynecol. 55 (3) (2016) 405–409.
[34] C.S. Marcus, G.L. Maxwell, K.M. Darcy, C.A. Hamilton, W.P. McGuire, Current approaches and challenges in managing and monitoring treatment response in ovarian cancer, J. Cancer 5 (1) (2014) 25–30.
[35] G. Gossner, R.L. Coleman, D.G. Mutch, N.S. Horowitz, J.S. Rader, R.K. Gibb, M. A. Powell, T.J. Herzog, CA-125 response in patients with recurrent ovarian or primary peritoneal cancer treated with pegylated liposomal doxorubicin or topotecan, Gynecol. Oncol. 103 (1) (2006) 212–218.
[36] M. Schwartz, O. Camacho-Vanegas, A.M. Wood, M. Dashkoff, C. Whitelock, T. T. Harkins, C.J. Cohen, A.M. Beddoe, P. Dottino, J.A.

Martignetti, Applying precision medicine to ovarian Cancer: proof-of-Principle for a "Molecular second look, Int. J. Gynecol. Cancer 28 (3) (2018) 479–485.
[37] P. Lohmann, N. Galldiks, M. Kocher, A. Heinzel, C.P. Filss, C. Stegmayr, F.
M. Mottaghy, G.R. Fink, N. Jon Shah, K.J. Langen, Radiomics in neuro-oncology: basics, workflow, and applications, Methods (2020).
[38] K.J. Park, J.L. Lee, S.K. Yoon, C. Heo, B.W. Park, J.K. Kim, Radiomics-based prediction model for outcomes of PD-1/PD-L1 immunotherapy in metastatic urothelial carcinoma, Eur. Radiol. (2020).
[39] M.K. Santos, J.R. Ferreira Júnior, D.T. Wada, A.P.M. Tenorio, M.H.N. Barbosa,´ Marques PMA: artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine, Radiol. Bras.
52 (6) (2019) 387–396.
[40] L. Ma, Z. Xiao, K. Li, S. Li, J. Li, Yi X, Game theoretic interpretability for learning based preoperative gliomas grading, Future Gener. Comput. Syst. 112 (2020) 1–10.